AtomNet

A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery

Izhar Wallach, Michael Dzamba, Abraham Heifets

Victor Storchan, Institute for Computational and Mathematical Engineering

Table of contents

What are CNNs?

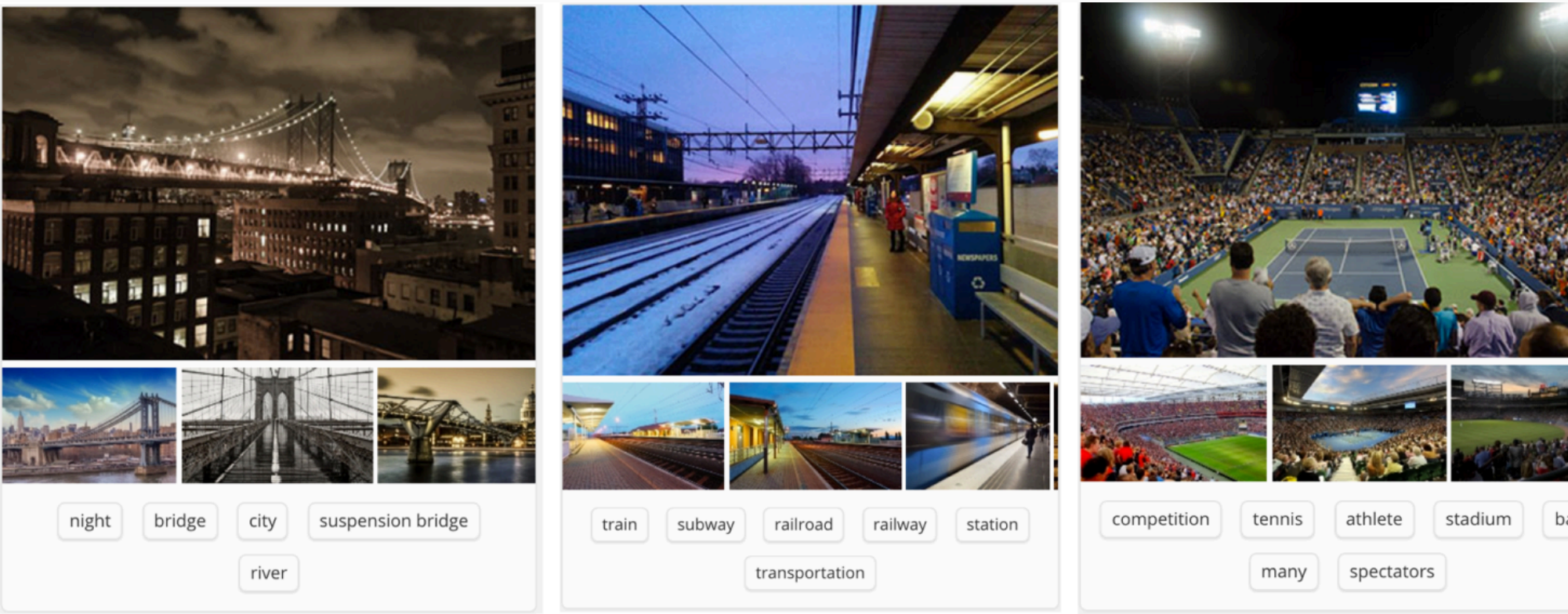


Figure 1: recognizing scenes and the system is able to suggest relevant tags

What are CNNs?

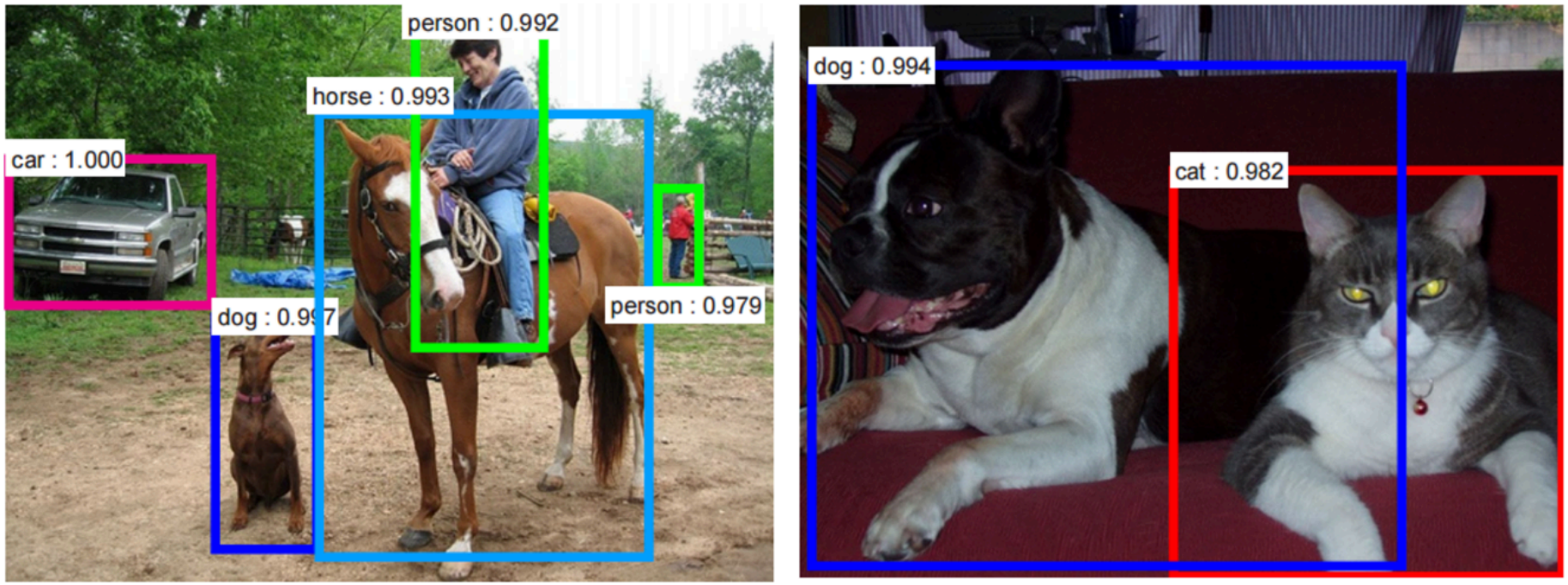


Figure 2: recognizing everyday objects, humans and animals

What are CNNs?

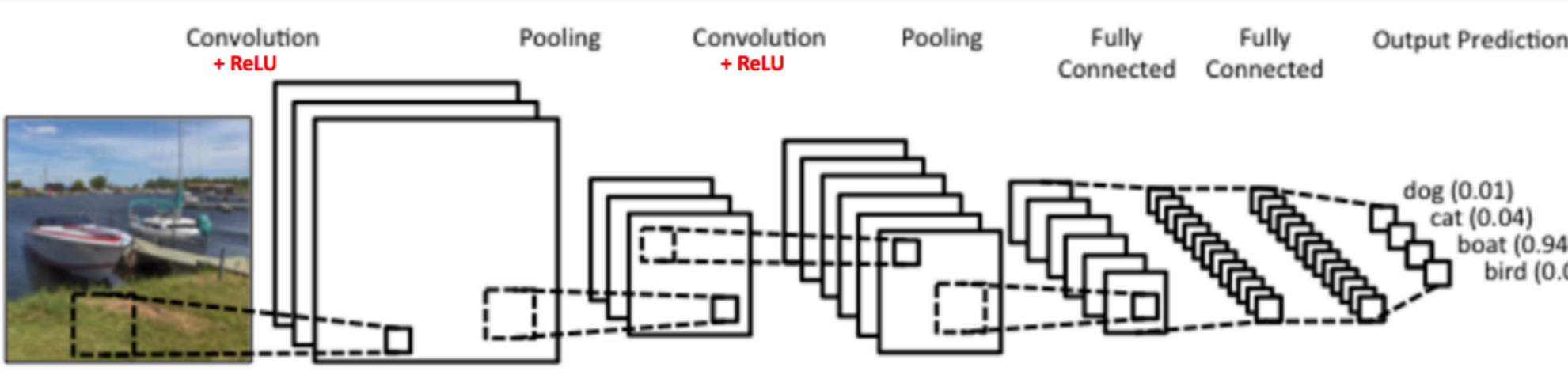


Figure 3: Architecture for classification

Primary purpose of **Convolution**:

- **Extracting** features from the input image while preserving spatial relationship between pixels.
- **Learning** image features using small squares of input data.

Convolutional Neural Network for virtual screening

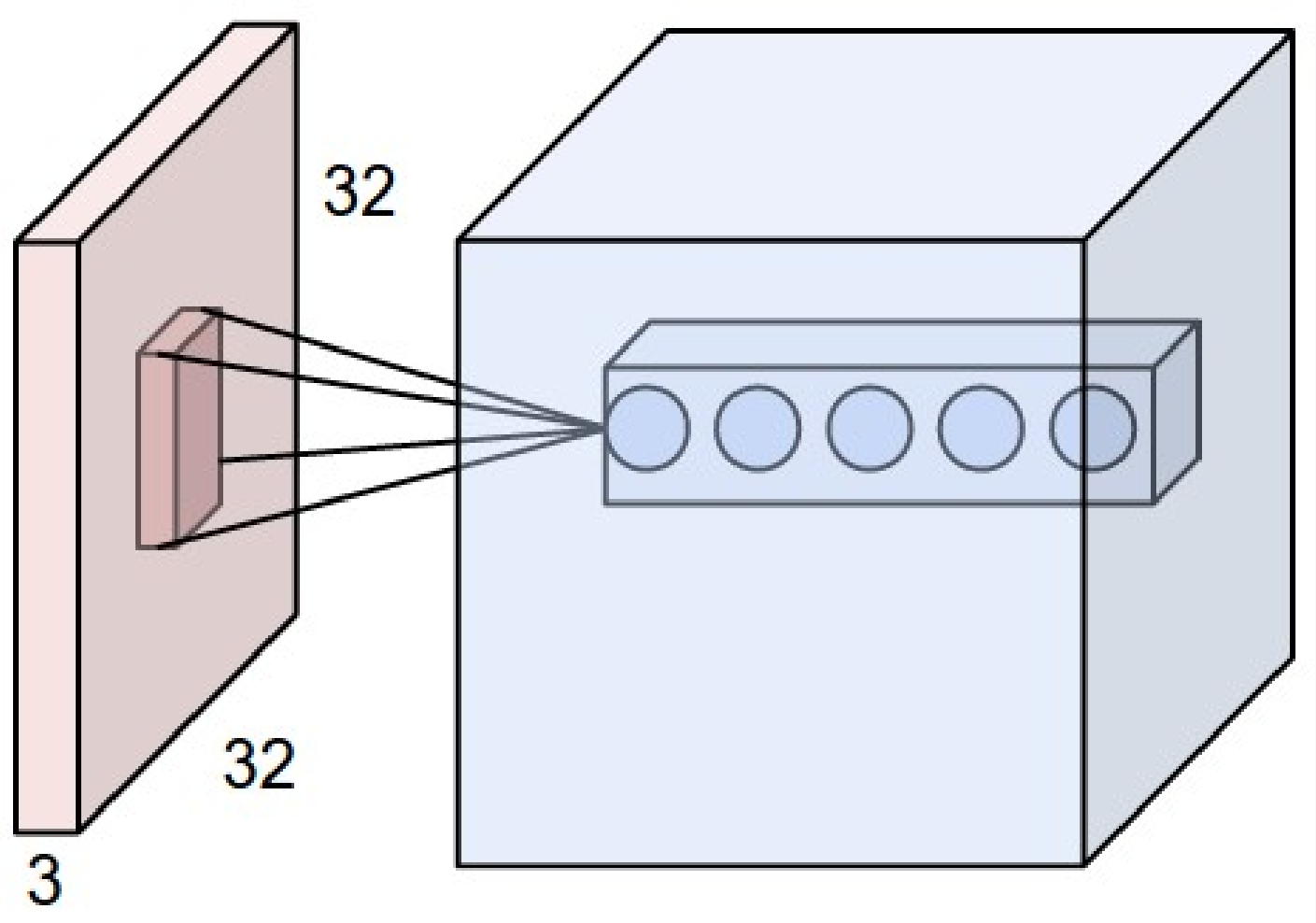


Figure 4: From CS231n, use of 5 different filters on a 32*32*3 input

We want to learn the weights of the filters to discover features.

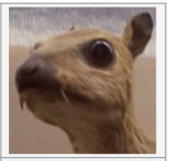
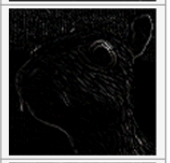
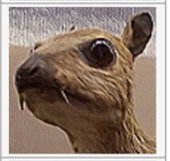
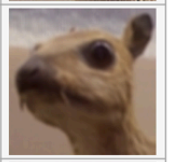
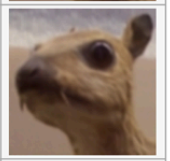
Operation	Filter	Convolved Image
Identity	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	
Edge detection	$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$	
	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	
	$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$	
Sharpen	$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$	
Box blur (normalized)	$\frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$	
Gaussian blur (approximation)	$\frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$	

Figure 5: Filters

Goal

Goal

Use of Deep Convolutional Neural Networks (CNN) to **predict the bioactivity of small molecules** for drug discovery application.

- Apply local convolutional filters to structural target input information.
- Biochemical interactions are local by nature so should be well-handled by these sparse ML architectures.
- Predict new active molecules for targets.

.

Drugs Design techniques

Ligand-based Drug Design

Given a set of diverse ligands that bind to a receptor, the goal is to create a model of the receptor with this information. It ends up searching for molecules with shape similar to the ones of known activities.

Advantages:

- Computationally very efficient

Drawbacks:

- Prevent serendipity in drug discovery

Structured-based Drug design

It involves docking of candidate ligands into a protein target + scoring function (based on energy) to estimate the likelihood that the ligand will bind to the protein.

Convolutional Neural Network for virtual screening

Convolutional Neural Network: Architecture

Network architecture 3D convolutional layers were implemented to support parameters such as filter size, stride, and padding in a similar fashion to the implementation of Krizhevsky *et al.* [4]. We used network architecture of an input layer as described above, followed by four convolutional layers of 128×5^3, 256×3^3, 256×3^3, 256×3^3 (number of filters $\times$ filer-dimension), and two fully-connected layers with 1024 hidden units each, topped by a logistic-regression cost layer over two activity classes.

Model Training Training the model was done using stochastic gradient descent with the AdaDelta adaptive learning method [34], the backpropagation algorithm [35], and mini-batches of 768 examples per gradient step. No attempt was made to optimize meta-parameters except the limitation of fitting the model into a GPU memory. Training time was about a week on 6 Nvidia-K10 GPUs.

Typically, the **cross-entropy loss function** is used for trqining a probabilistic classification:

$$L(f,(x,y)) = -ylog(f(x)) - (1-y)log(1-f(x)), \text{ for } y \text{ binary.}$$

Method

Input representation

The input representation is a 5 steps process which results in a 1D vector that is given to the input layer.

1. **Find the binding site** with a flooding algorithm (exploration through the surface of the protein)
2. **Define a Cartesian 3D grid** which center is set to be the center-of-mass of the binding site.
3. **Choose a pose** within the binding site cavity.
4. **Crop the geometric data** to fit within an appropriate bounding box.
5. **Unfold the 3D grid** into a 1D floating point vector.

Values: Enumeration of atom types.

Datasets

1. **The Directory of Useful Decoys Enhanced (DUDE):**
   - **Gather diverse sets of active molecules** for a set of target proteins.
   - **Prevent analogue bias**: cluster according to similar active. Take one representative of each class.
   - Each active comes with a set of property matched decoys (PMD, inactive).
   - **Benchmark:** 102 targets, 22,886 actives (average of 224 actives per target) + 50 PMD per active.
   - **Test set:** 30 targets
   - **Training set:** remaining 72 targets
2. **ChEMBL-20 PMD:** another DUDE-like dataset.

Datasets: Limitations

No experimental validation to verify that decoys are actually inactive. Decoys are chosen topologically very dissimilar from the actives.

The method is blind to cases where shapes of decoys are close to the ones of actives.

Results

Baseline: Smina

Smina is used as the baseline. It implements an improved empirical scoring function and minimization.

AUC		> 0.5	> 0.6	> 0.7	> 0.8	> 0.9
ChEMBL-20 PMD	AtomNet	49	44	36	24	10
	Smina	38	10	4	1	0
DUDE-30	AtomNet	30	29	27	22	14
	Smina	29	25	14	5	1
DUDE-102	AtomNet	102	101	99	88	59
	Smina	96	84	53	17	1
ChEMBL-20 inactives	AtomNet	149	136	105	45	10
	Smina	129	81	31	4	0

Table 2: The number of targets on which AtomNet and Smina exceed given AUC thresholds. For example, on the CHEMBL-20 PMD set, AtomNet achieves an AUC of 0.8 or better for 24 targets (out of 50 possible targets). ChEMBL-20 PMD contains 50 targets, DUDE-30 contains 30 targets, DUDE-102 contains 102 targets, and ChEMBL-20 inactives contains 149 targets.

Figure 6:

Conclusion

Filters specializations

Apply filter to input data and map the location of biggest magnitude to the relevant subset of the site of the binding. This subset has specific chemical functions and the filter is specializing as a detector of these functions.

Remarks

1. We do not know anything about any validation step to set the number of hidden layers.
2. We do not know about how the initialization of the weights is done (usually, it is drawn from a uniform distribution)
3. Why using backpropagation instead of a second order method like conjugate gradient which could allow for faster convergence?

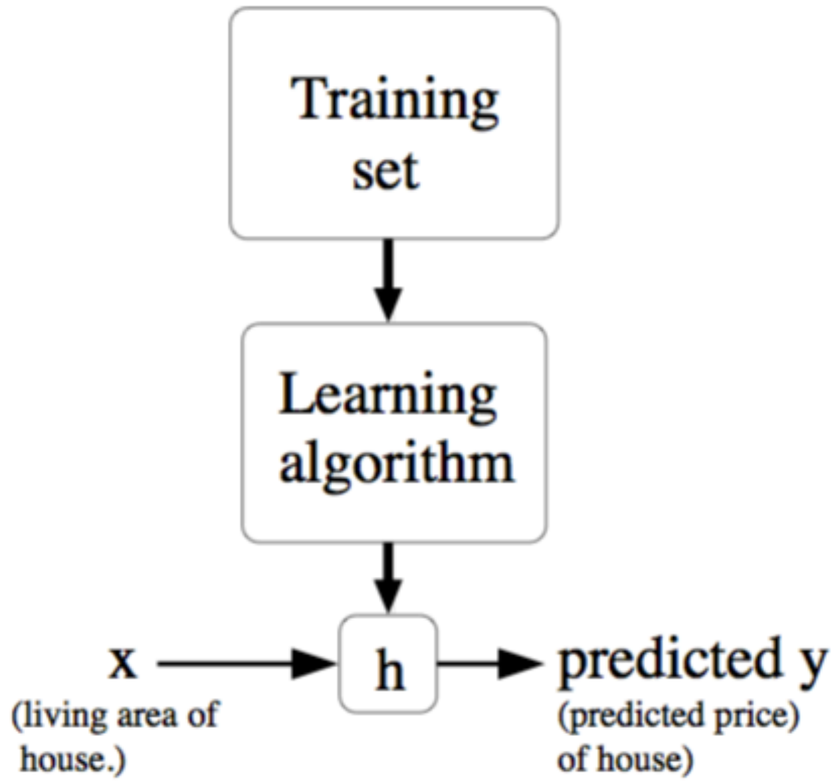
Introduction to Machine Learning and Deep Learning

Lawrence Lin Murata

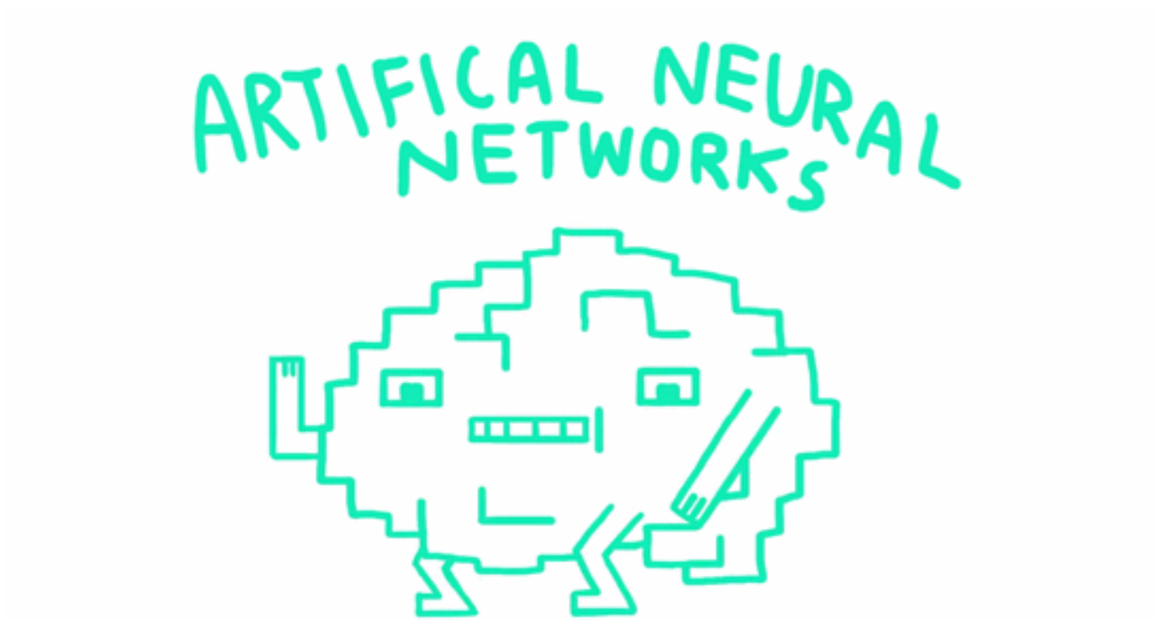
(Very Quick) Introduction to Machine Learning and Deep Learning

Lawrence Lin Murata

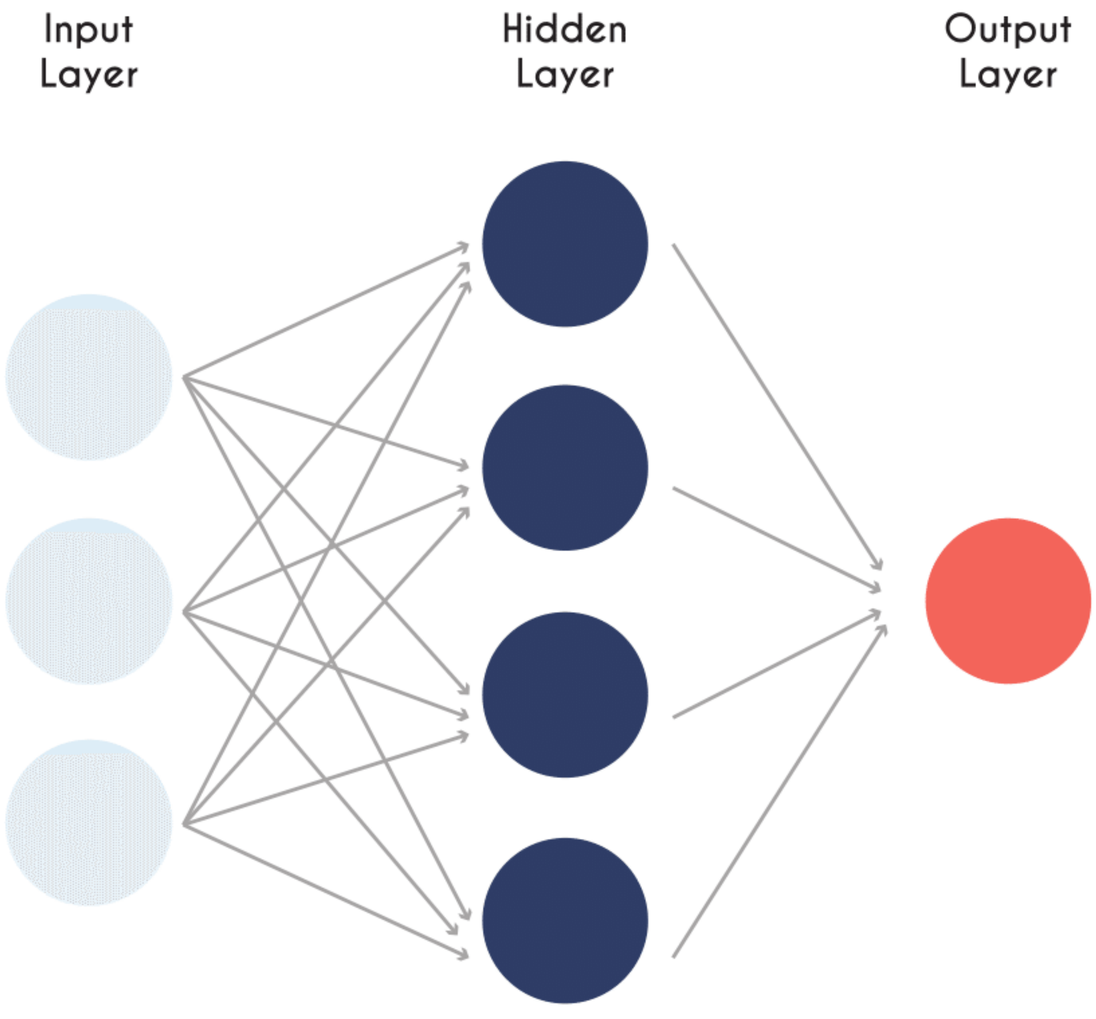
Machine Learning



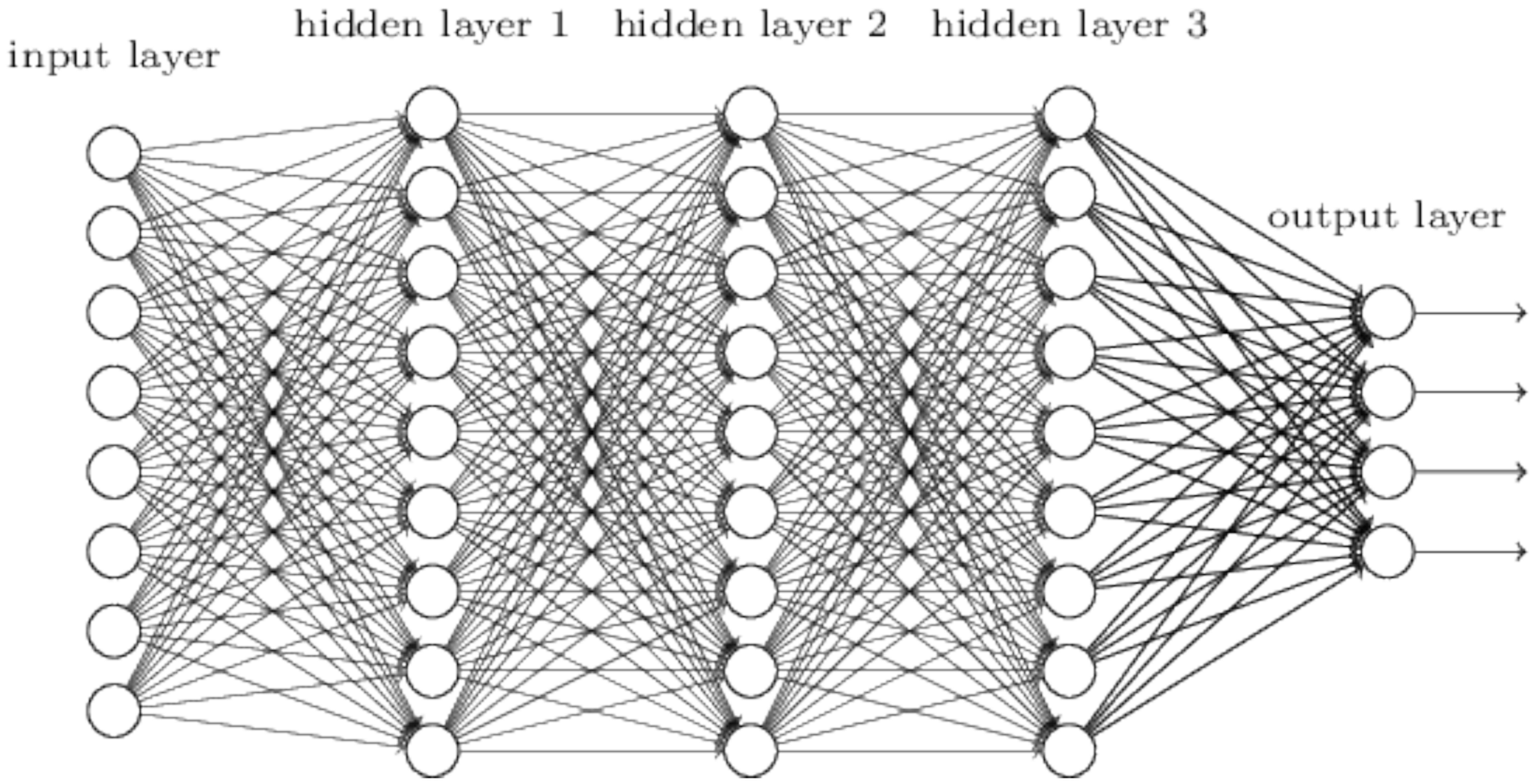
Neural Networks



Neural Networks



Deep Learning



Learning Deep Architectures for Interaction Prediction in Structure-based Virtual Screening

Lawrence Lin Murata

Learning Deep Architectures for Interaction Prediction in Structure-based Virtual Screening

Learning Deep Architectures

Learning Deep Architectures for Interaction Prediction in Structure-based Virtual Screening

for Interaction
Prediction

for Interaction
Prediction

for Interaction
Prediction

+

Learning Deep Architectures for Interaction Prediction in Structure-based Virtual Screening

in Structure-based Virtual Screening

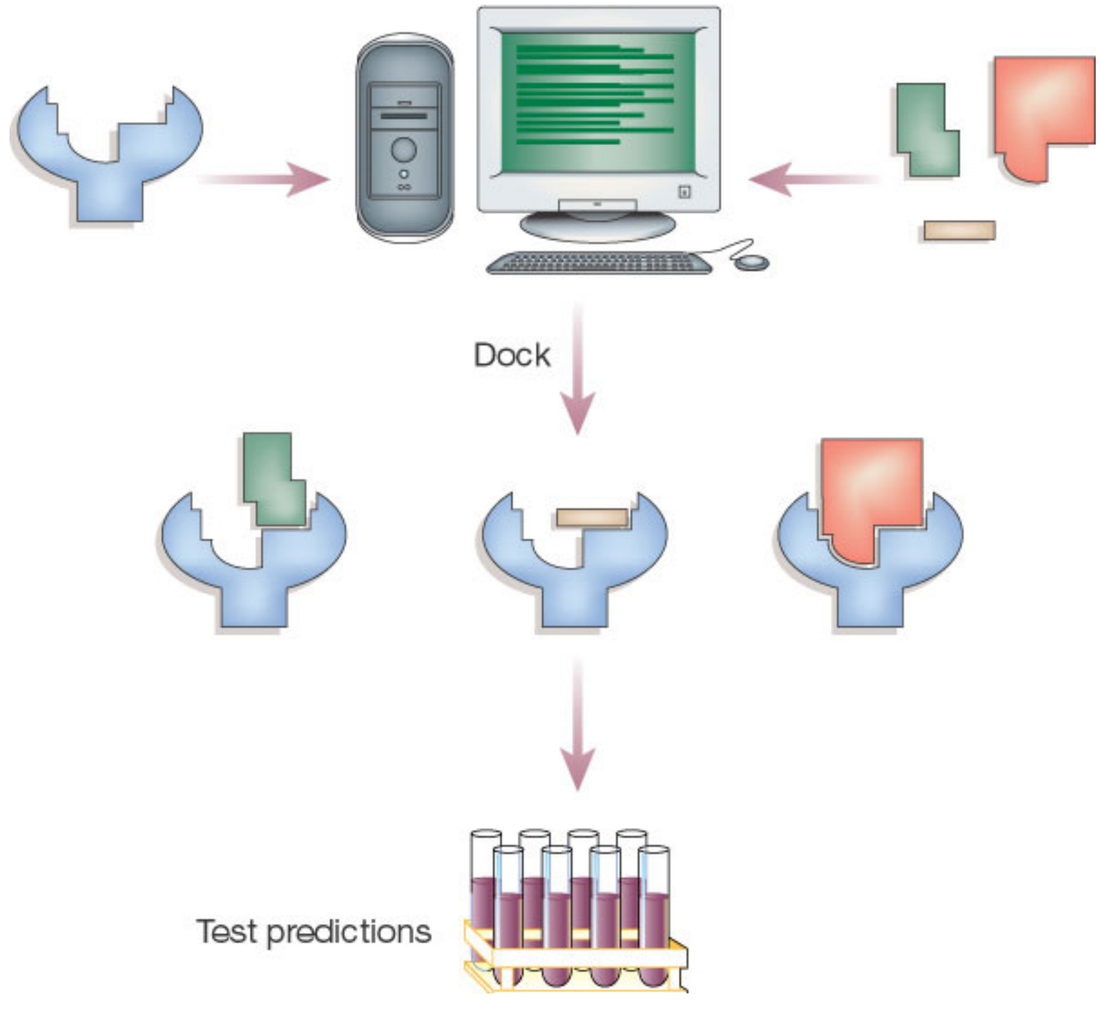
in Structure-based
Virtual Screening

STRUCTURE VS. **LIGAND**

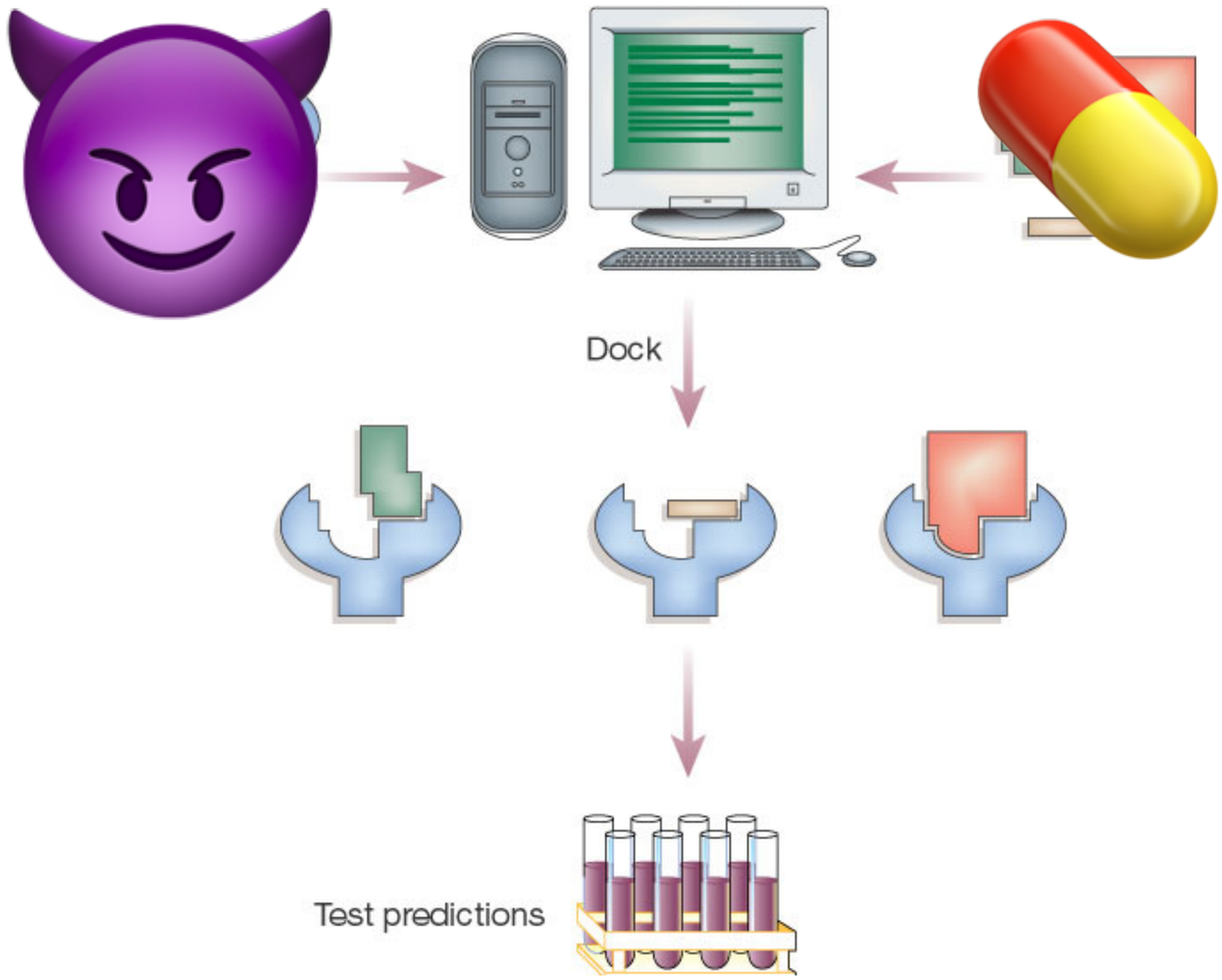
in Structure-based
Virtual Screening

STRUCTURE

in Structure-based Virtual Screening

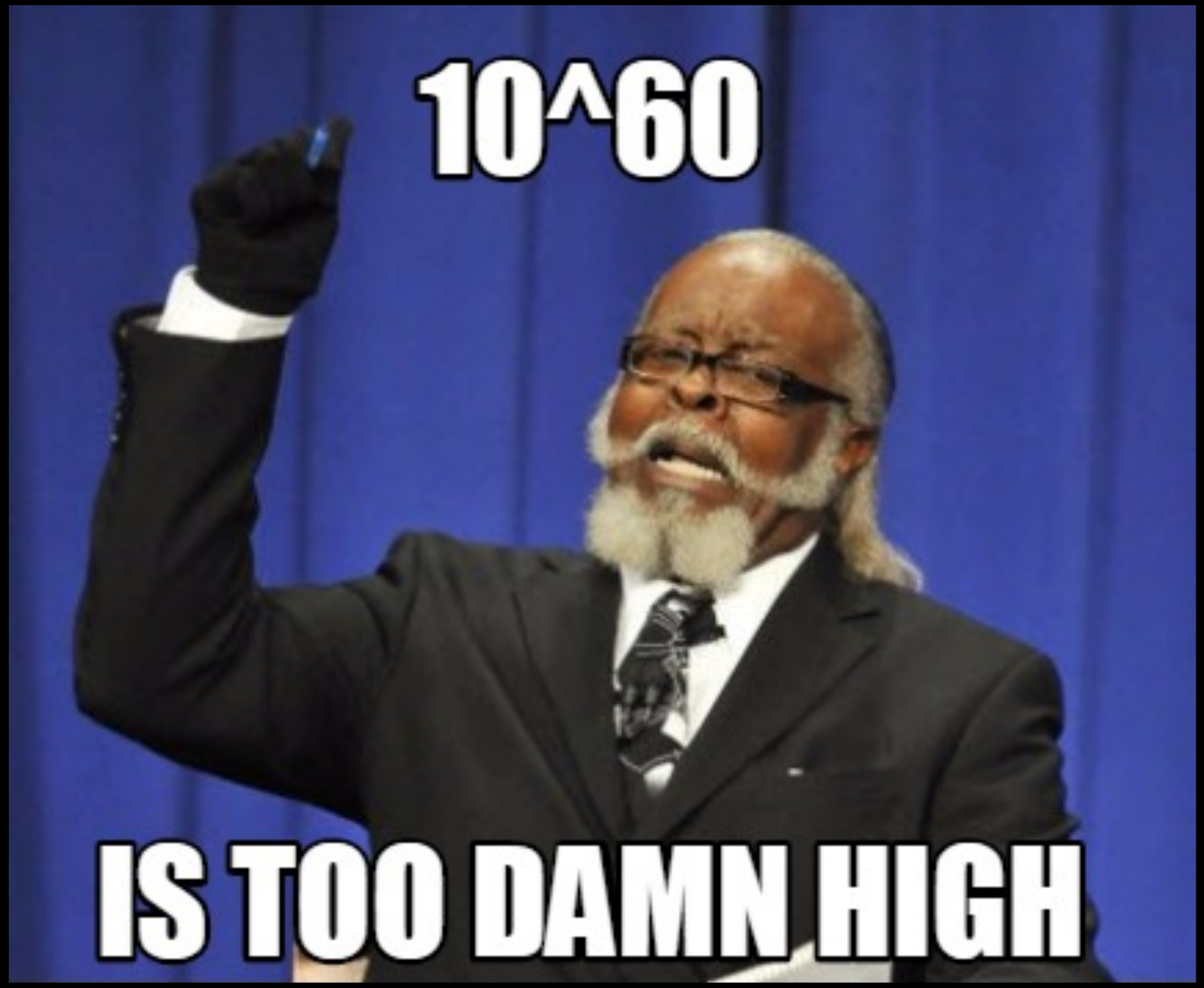


in Structure-based Virtual Screening



Problems in Structure-based Virtual Screening

1,000,000,000,000,000,000
,000,000,000,000,000,000,
000,000,000,000,000,000,0
00,000,000,000,000,000


10^60
IS TOO DAMN HIGH

Problems in Structure-based Virtual Screening

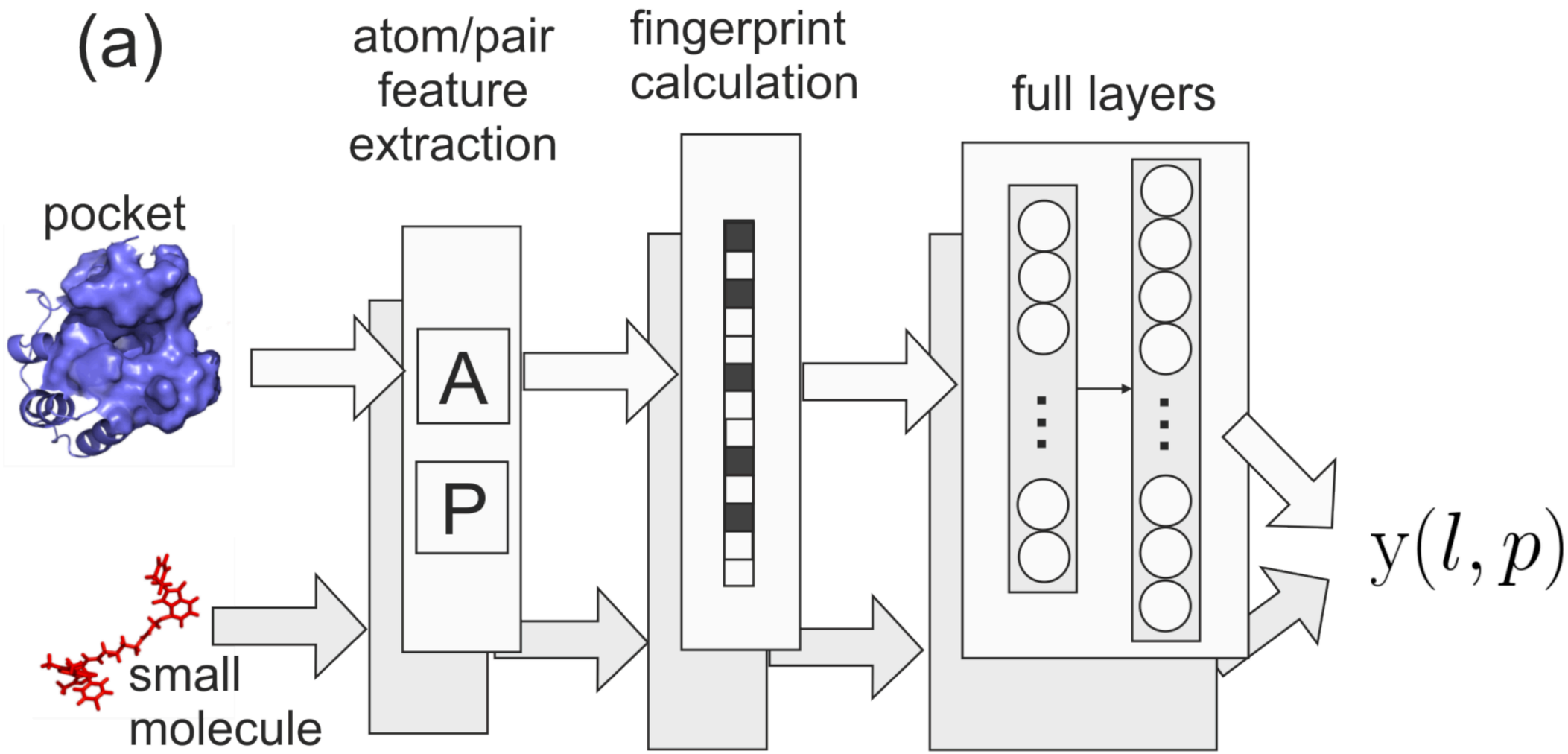
1. Complex chemical space

2. Lack of exhaustive training data

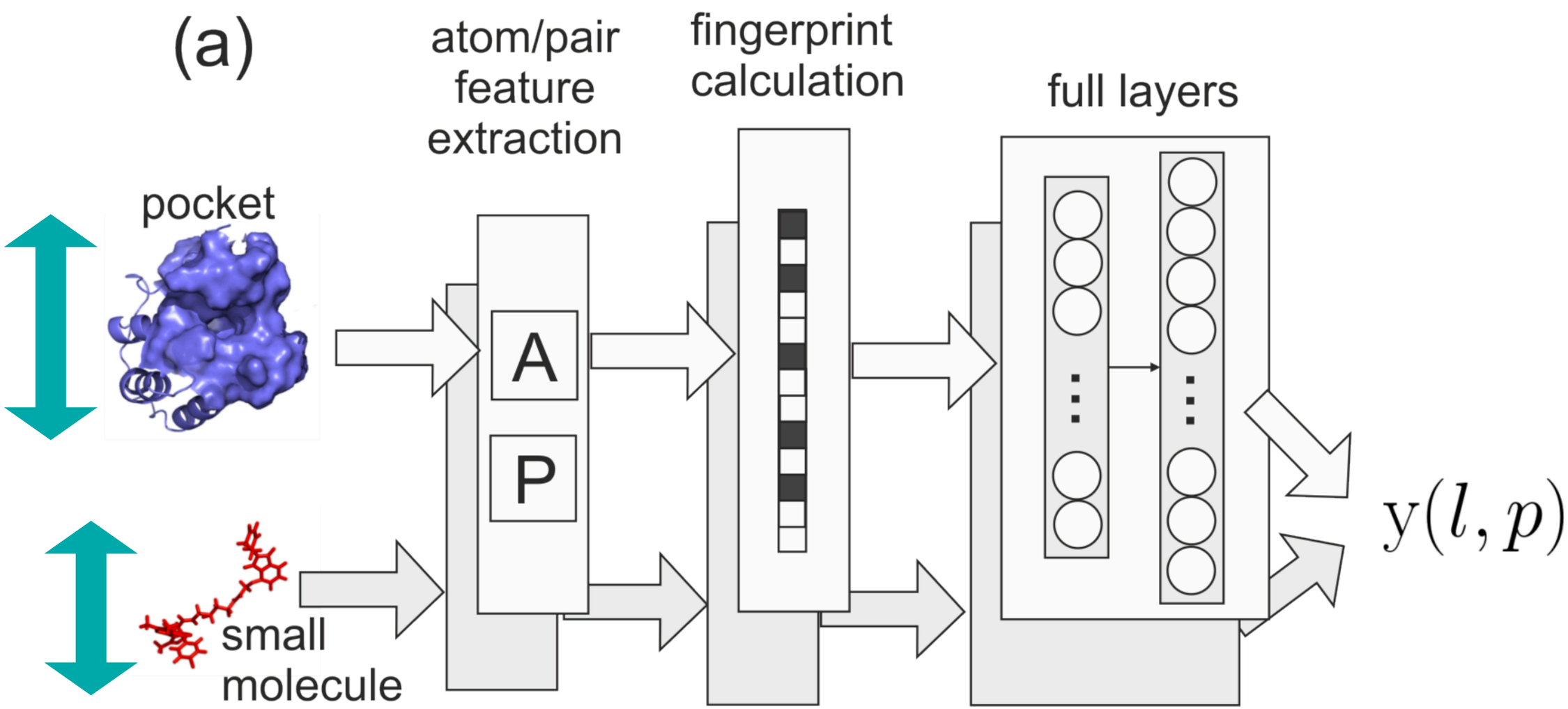
3. High number of false positives

Methods

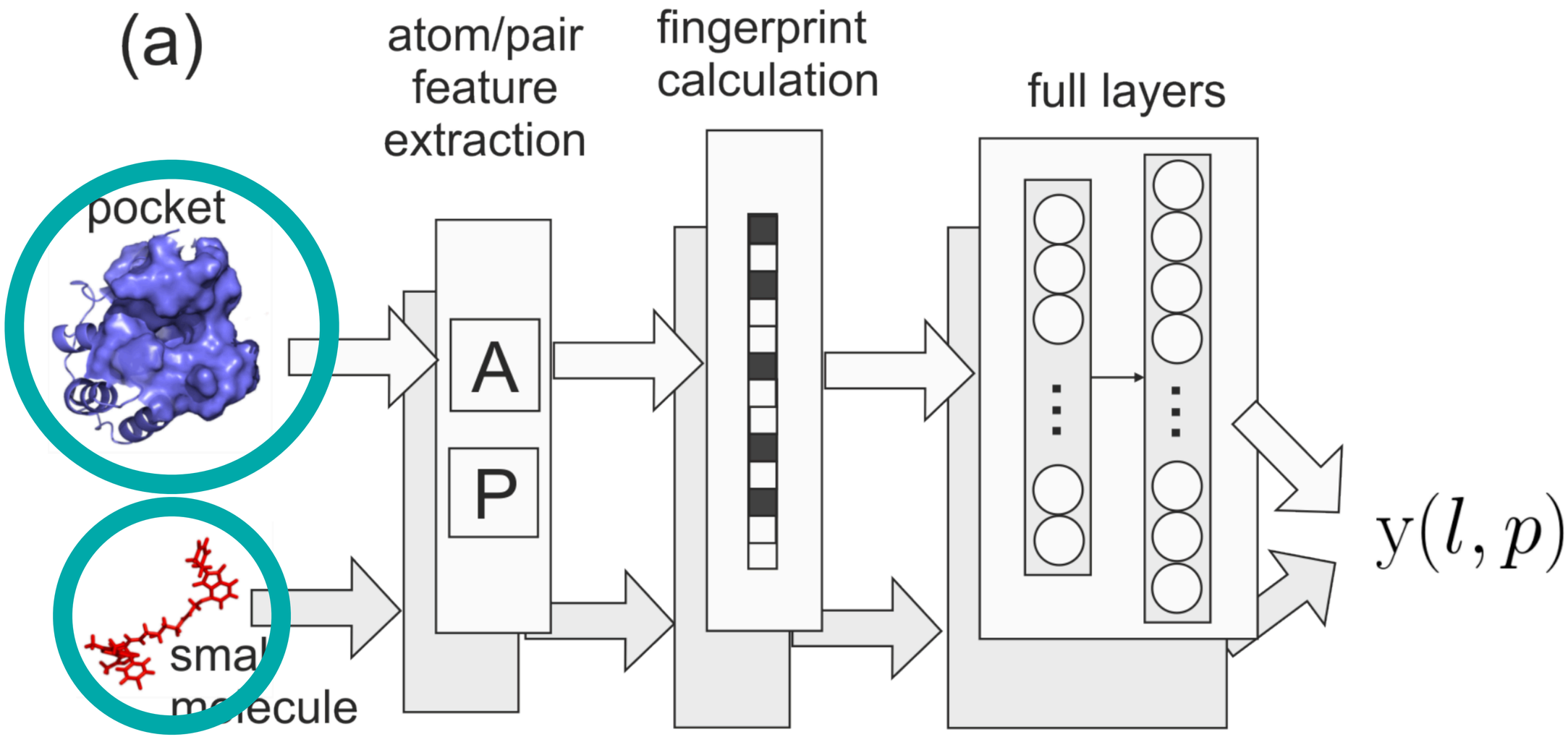
Method


(a)
atom/pair
feature
extraction
fingerprint
calculation
full layers
pocket
A
P
small
molecule
y(l, p)

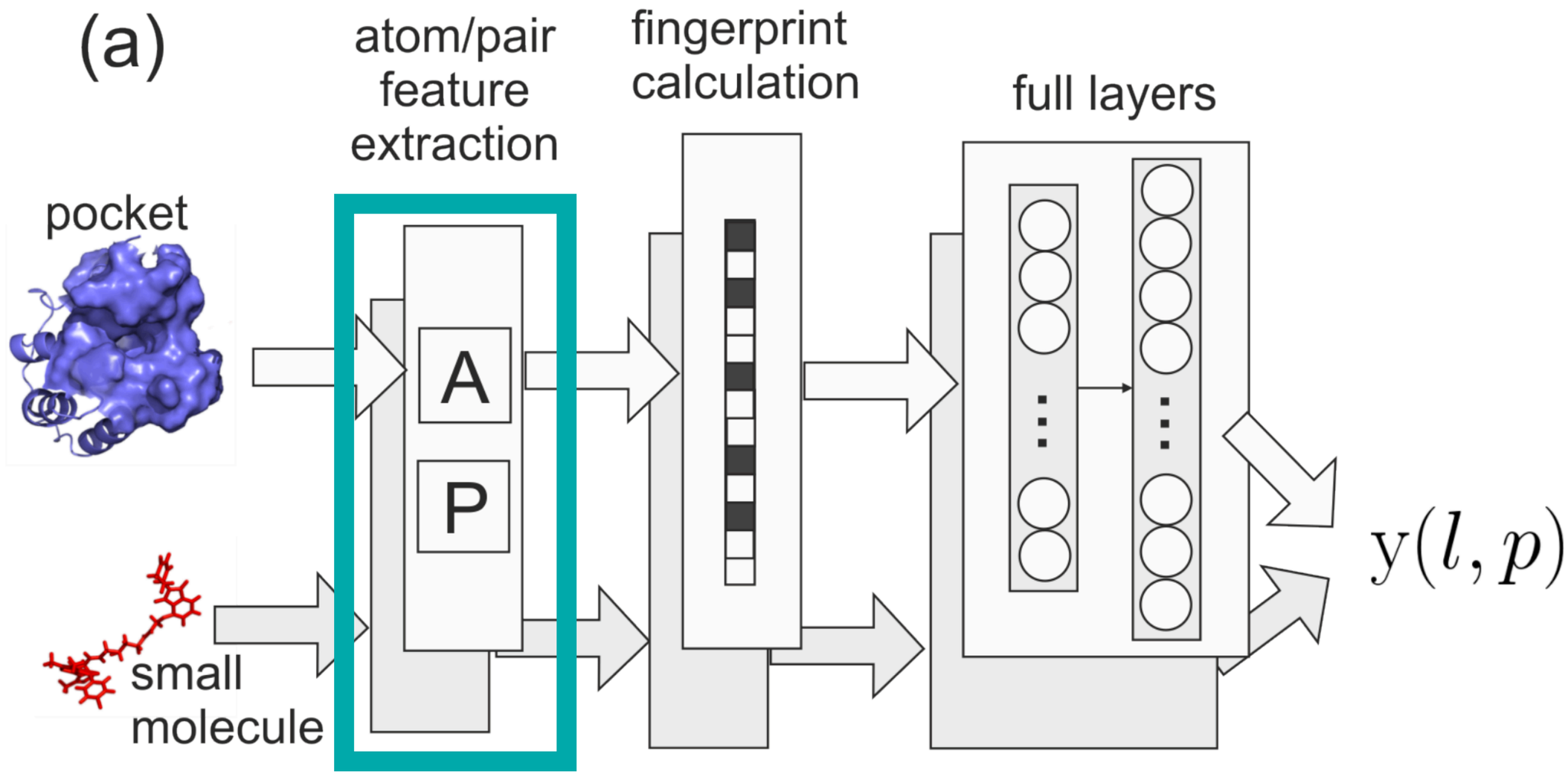
Challenges


(a)
atom/pair feature extraction
fingerprint calculation
full layers
pocket
A
P
small molecule
$y(l, p)$

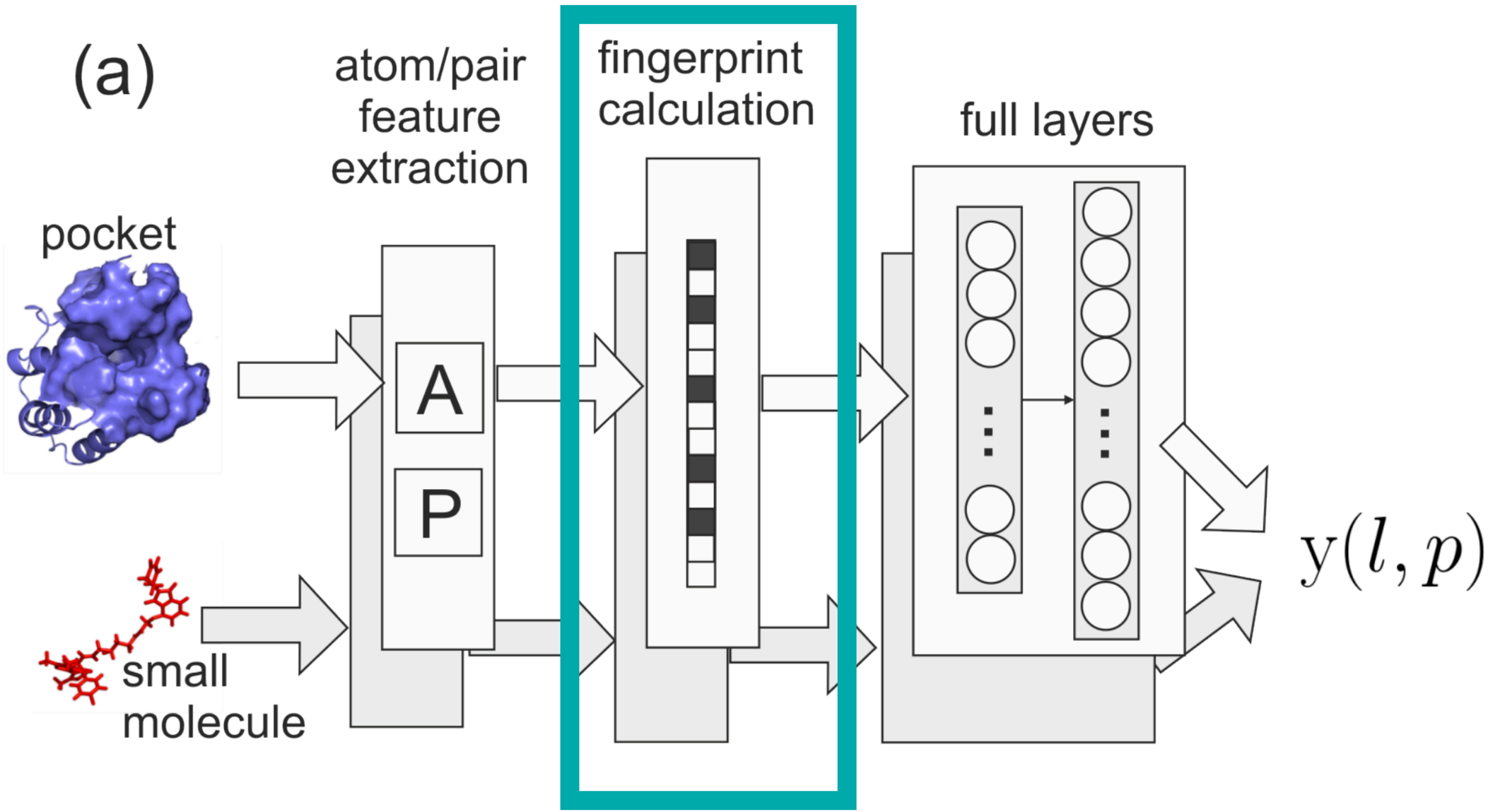
Challenges


(a)
atom/pair feature extraction
fingerprint calculation
full layers
pocket
A
P
small molecule
$y(l, p)$

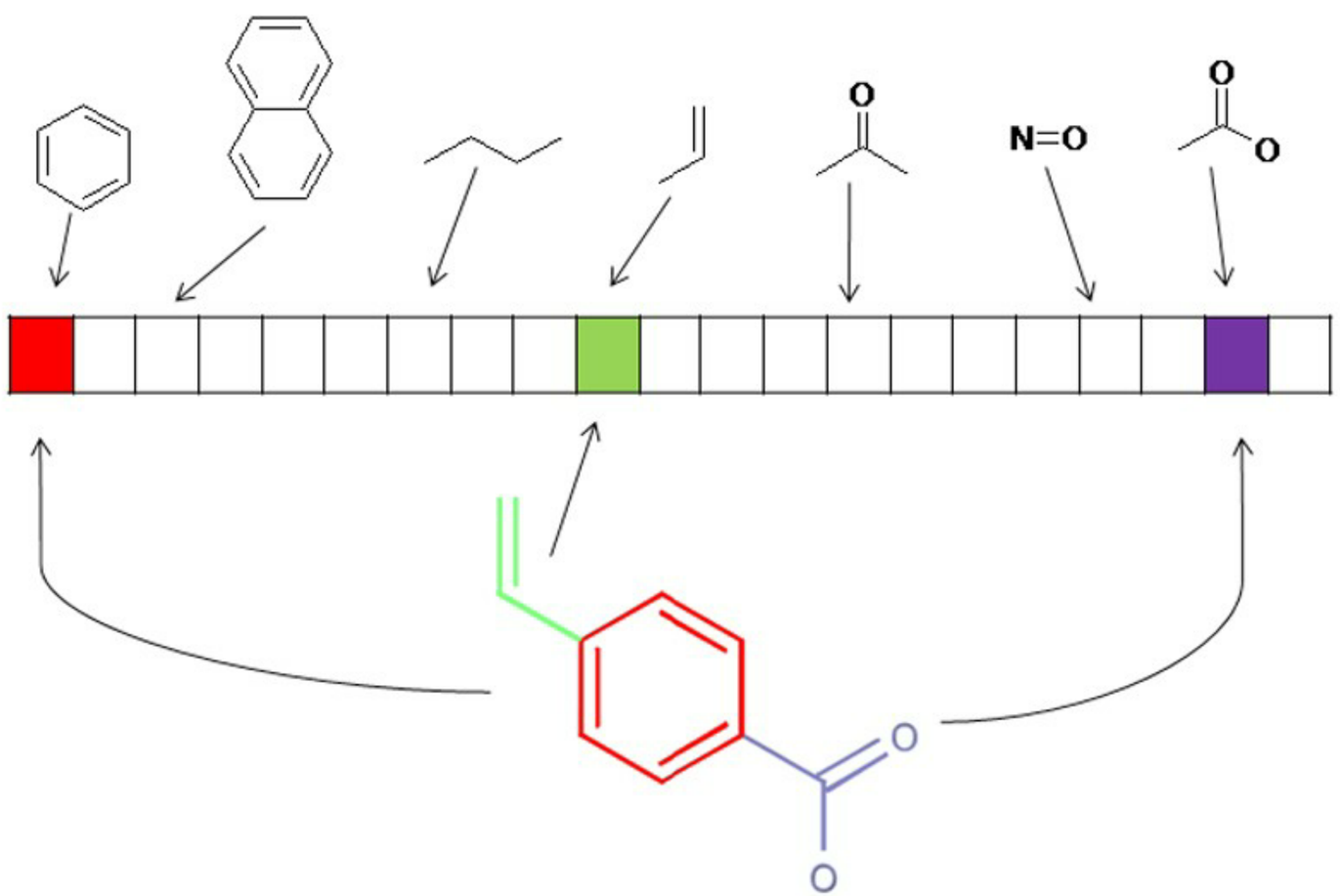
Challenges


(a)
atom/pair
feature
extraction
fingerprint
calculation
full layers
pocket
A
P
small
molecule
$y(l, p)$

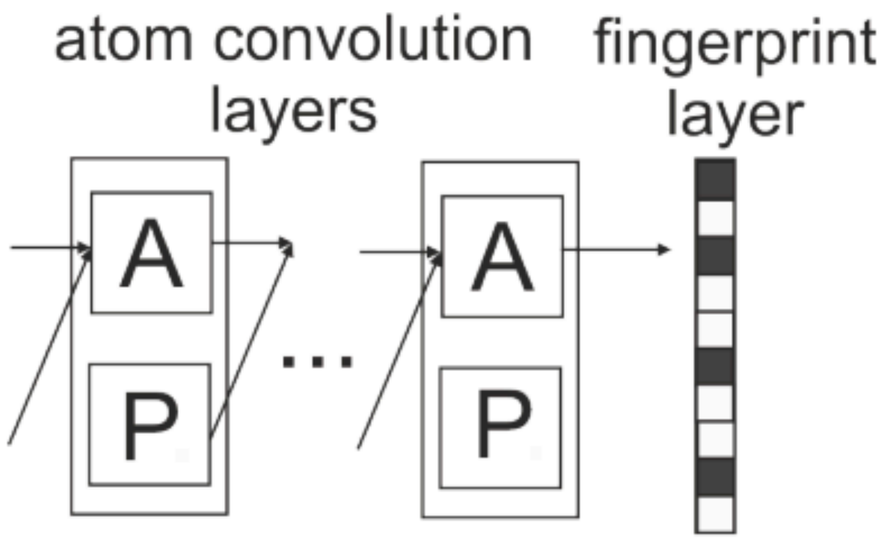
Fingerprint



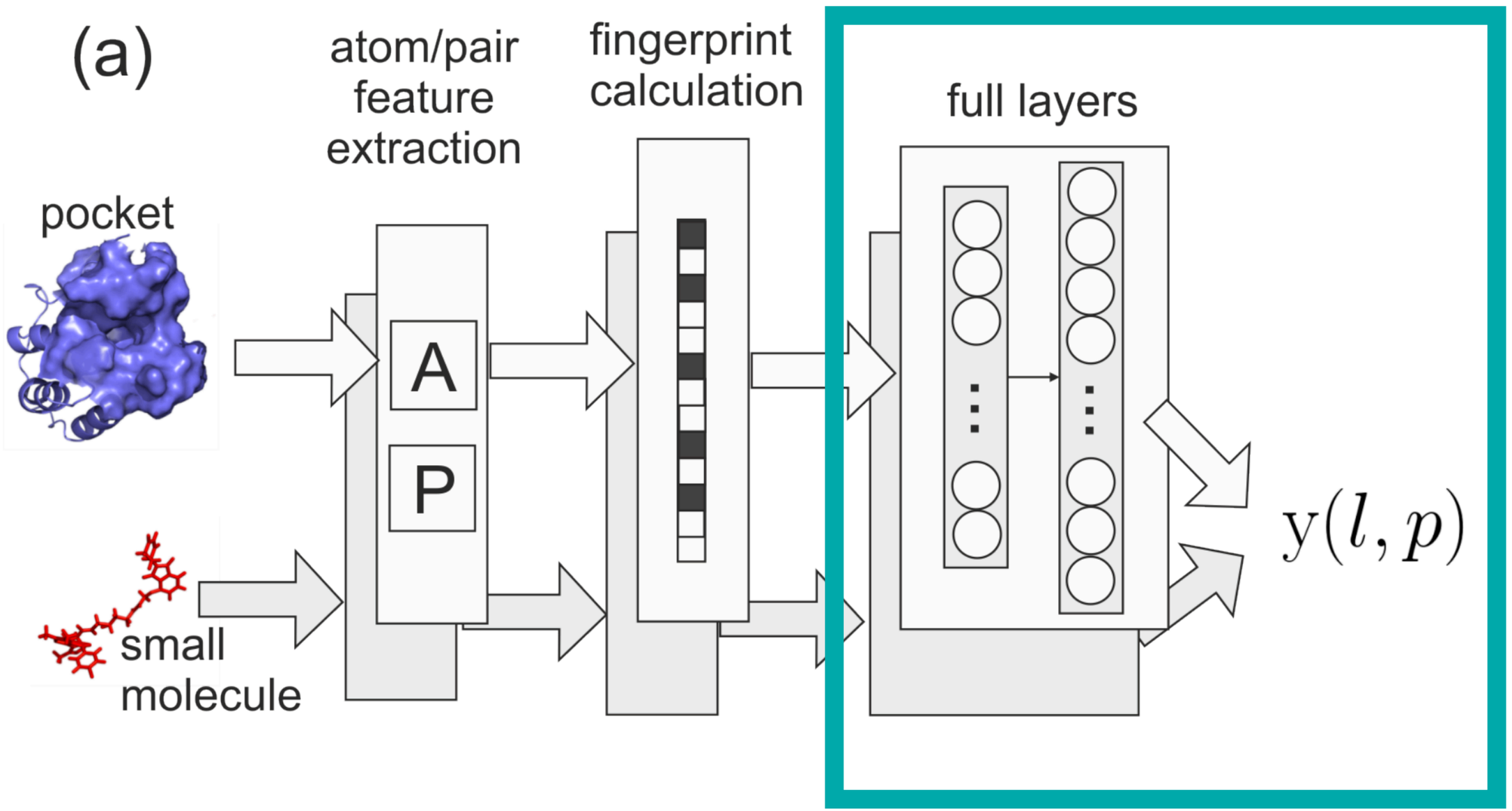
Fingerprint



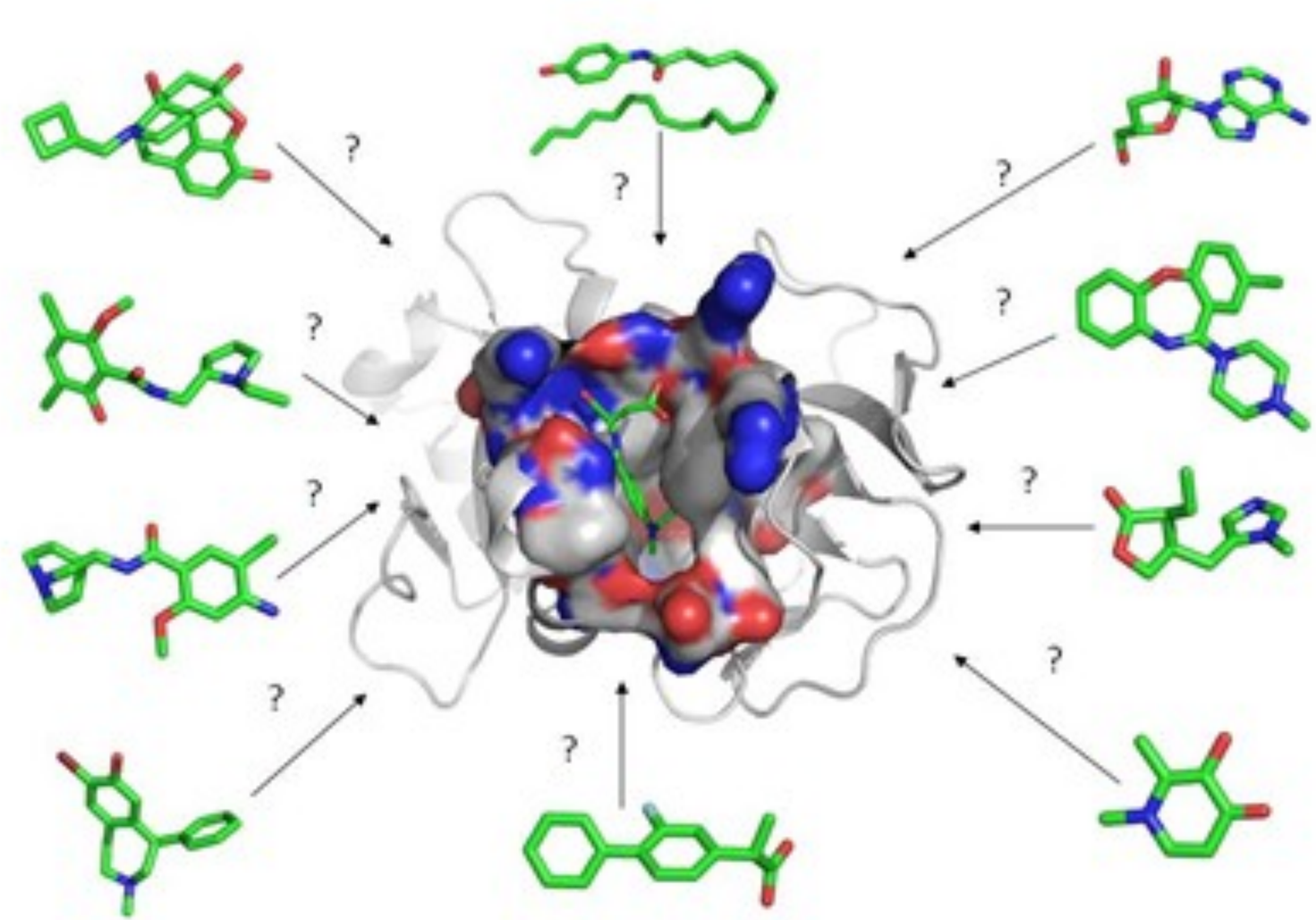
Neural Fingerprint



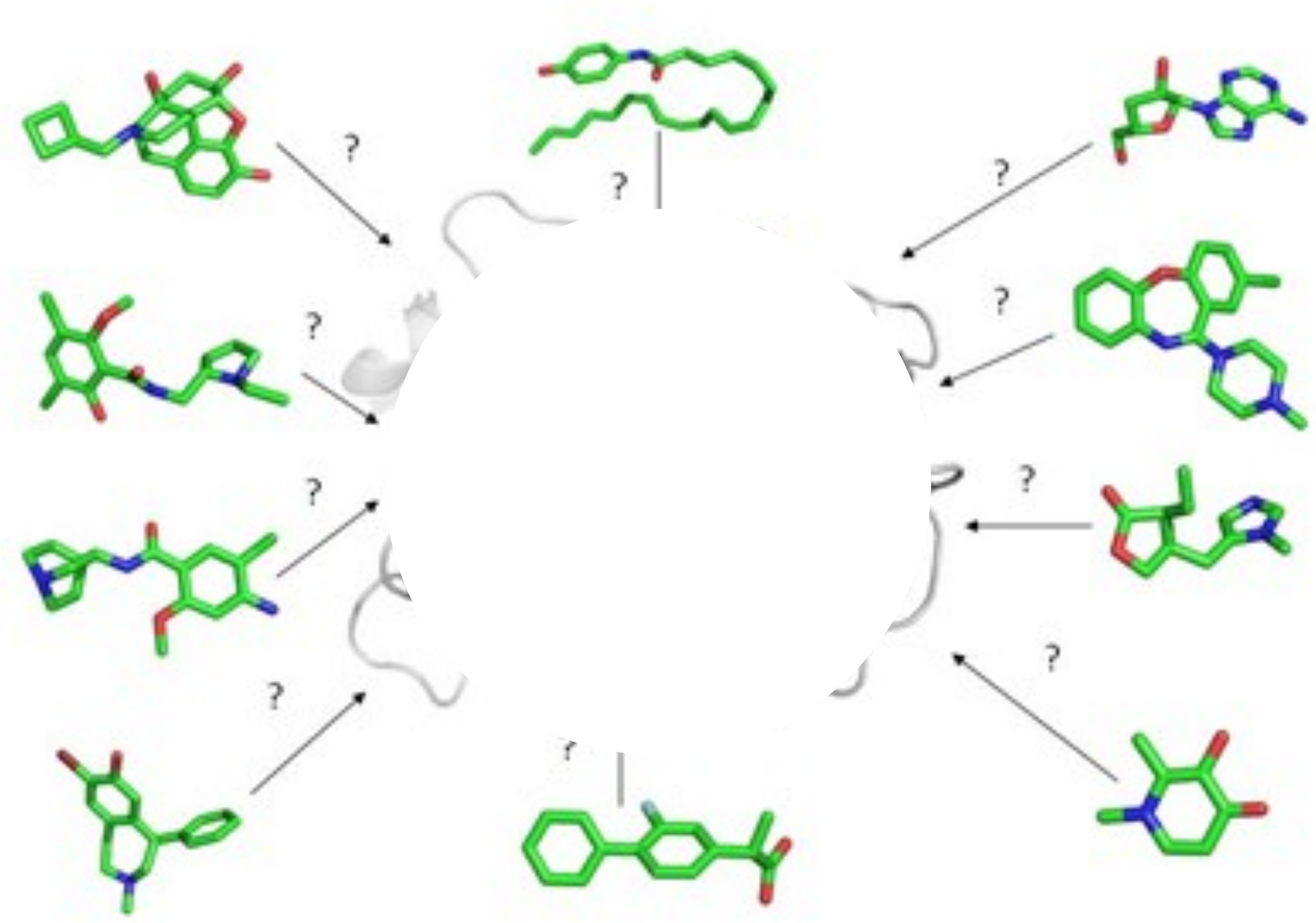
Method


(a)
atom/pair
feature
extraction
fingerprint
calculation
full layers
pocket
A
P
small
molecule
$y(l, p)$

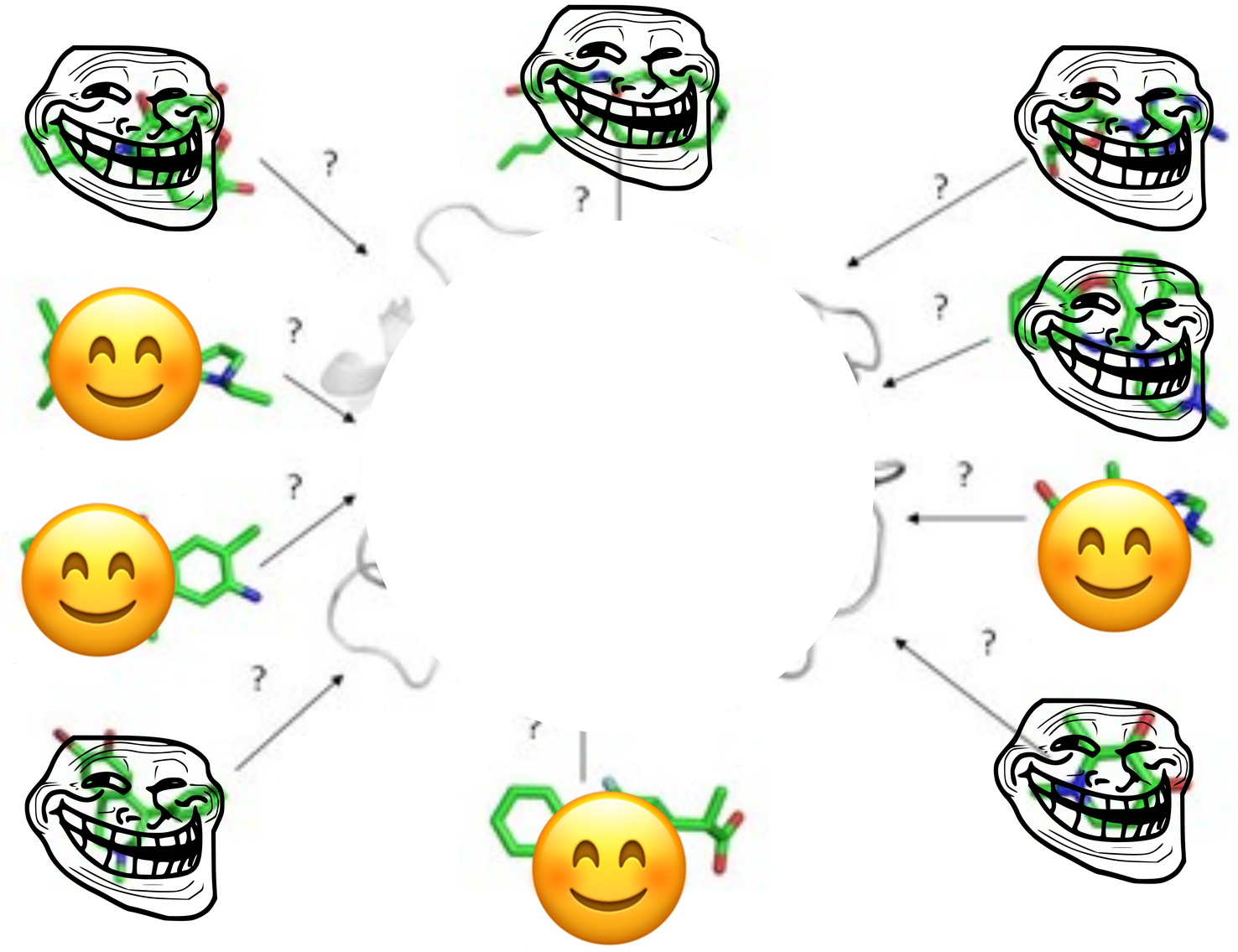
Method



Method



Method



Results

Results

Table 1: Results on DUD-E benchmark (70% of data for training and 30% of data for testing) and on DUD benchmark (leave-one-out cross-validation).

Dataset	Method	Mean AUC
DUD-E	Smina	0.700
	AtomNet [14]	0.855
	cmpds ECFP + LR	**0.904**
DUD	DeepVS [9]	0.800

Results

Table 1: Results on DUD-E benchmark (70% of data for training and 30% of data for testing) and on DUD benchmark (leave-one-out cross-validation).

Dataset	Method	Mean AUC
DUD-E	Smina	0.700
	AtomNet [14]	0.855
	cmpds ECFP + LR	**0.904**
DUD	DeepVS [9]	0.800

Results

Method	Total AUC	Mean AUC ($\pm$ std.)	AUC ≥ 0.7	AUC ≥ 0.8	AUC ≥ 0.9
AutoDock Vina	0.644	0.691 ± 0.147	47	21	4
Smina	0.653	0.704 ± 0.138	**54**	23	4
Ours(ECFP)	0.600	0.551 ± 0.166	21	2	0
Ours(NF)	**0.714**	$\mathbf{0.705 \pm 0.168}$	47	**29**	**11**

Future


CAN I HAZ
MOAR DATA?
memegenerator.net

Future

1. More training data

2. More accurate learnable fingerprints

3. Better learning techniques

4. More experimentation

5. Real impact

Machine - Learning Scoring Functions to Improve Structure - Based Binding Affinity Prediction And Virtual Screening

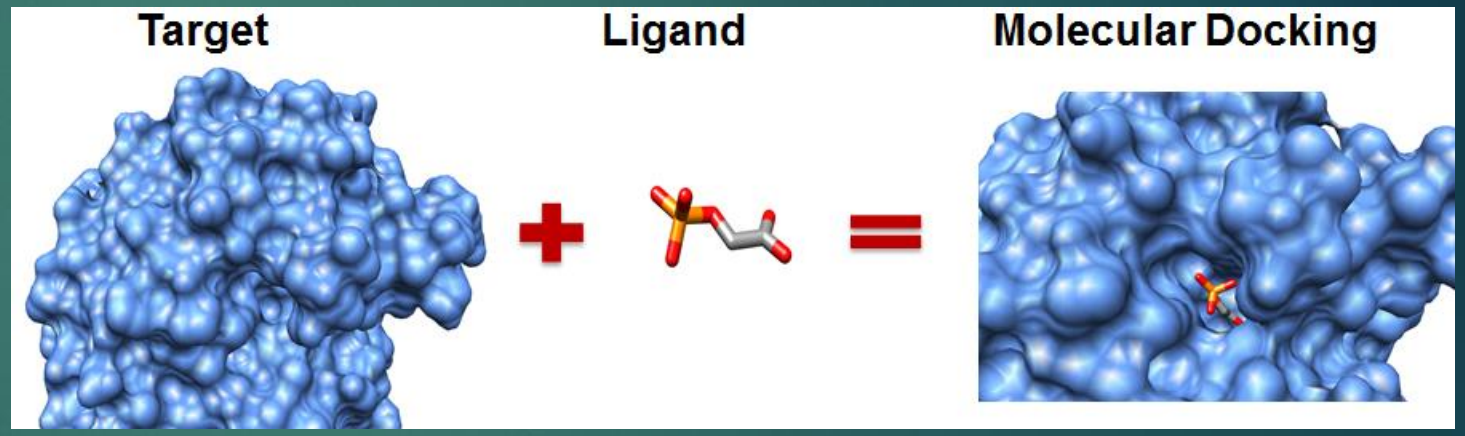
CHRISTOPHER VO

Overview

- Background
- Docking and Classical Scoring Functions
- Generic Machine Learning Scoring Functions for Binding Affinity
- Family Specific Machine Learning Scoring Functions
- Machine Learning Scoring Functions for Virtual Screening
- Emerging Applications of Machine Learning Scoring Functions

Background

- Docking
- Scoring Functions/Binding Affinity
- Virtual Screening



http://www.intechopen.com/books/protein-engineering-technology-and-application/protein-protein-and-protein-ligand-docking

Docking and Classical Scoring Functions

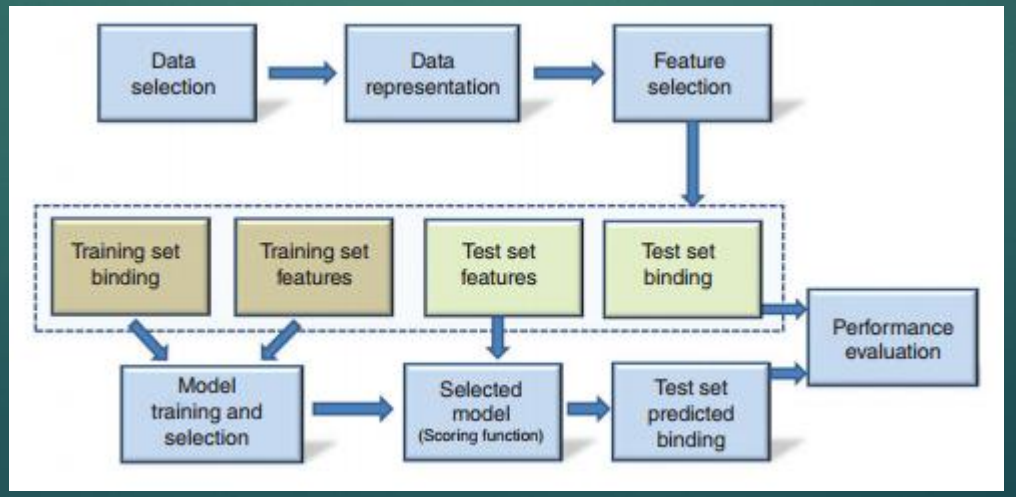
- Docking – two steps: pose generation, scoring
- Classical Scoring Functions and limitations

$$\Delta G = \Delta G_{VDW} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) + \Delta G_{hbond} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) +$$
$$\Delta G_{elec} \sum_{i,j} \frac{q_i q_j}{\epsilon (r_i j)^2} + \Delta G_{tor} N_{tor} + \Delta G_{sol} \sum_{i,j} (S_i V_j + S_j V_i) e^{\frac{-r_{ij}^2}{2\sigma^2}}$$

http://archive.cnx.org/contents/4e7287b0-6c38-4829-abeb-3ae357bbf60f@10/protein-ligand-docking-including-flexible-receptor-flexible-ligand-docking

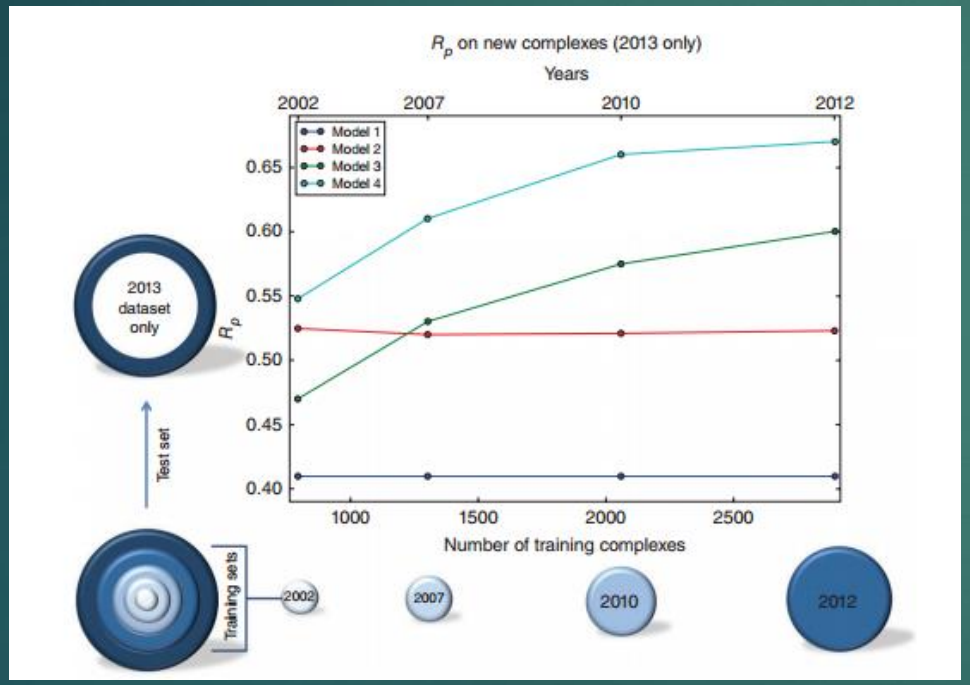
Machine Learning Scoring Functions

- Can capture more complex and non – linear characteristics
- Two applications: binding affinity and virtual screening
- Feature selection is very important
- Many benchmarks and metrics for performance (Pearson Correlation Coefficient for binding affinity, Enrichment Factor for virtual screening)



WIREs Comput Mol Sci 2015, 5:405–424. doi: 10.1002/wcms.1225

Generic Machine Learning Scoring Functions for Binding Affinity



- Work for many diverse protein – ligand complexes
- Earliest was Kernal-Partial Least Squares, showed nonlinear machine learning scoring function could capture functional form of binding affinity
- PDBbind benchmark standard for comparing performance

WIREs Comput Mol Sci 2015, 5:405–424. doi: 10.1002/wcms.1225

RF - Score

- Random Forest
- Features: number of protein – ligand atom type pairs within a certain range: C, N, O, F, P, S, Cl, Br, I
- Training (1105 complexes) and test (195 complexes) sets have no complexes in common
- Results
  - significantly better performance than 16 classical scoring functions
  - Performance increases with training set size unlike classical scoring functions

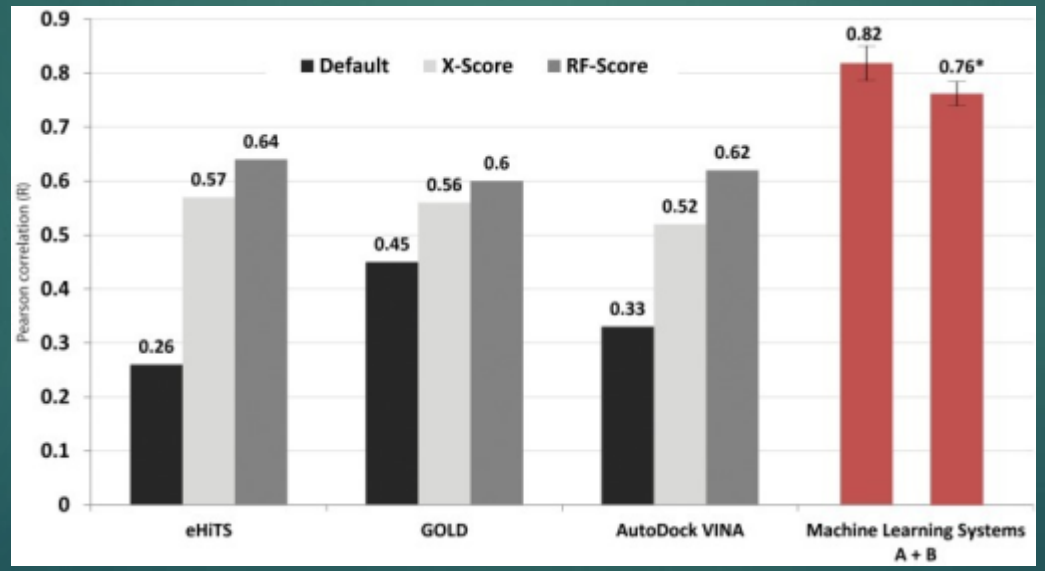
Table 2. Performance of scoring functions on the PDBbind benchmark

Scoring function	R	R_s	SD
RF-Score	0.776	0.762	1.58
X-Score::HMScore	0.644	0.705	1.83
DrugScore^{CSD}	0.569	0.627	1.96
SYBYL::ChemScore	0.555	0.585	1.98
DS::PLP1	0.545	0.588	2.00
GOLD::ASP	0.534	0.577	2.02
SYBYL::G-Score	0.492	0.536	2.08
DS::LUDI3	0.487	0.478	2.09
DS::LigScore2	0.464	0.507	2.12
GlideScore-XP	0.457	0.435	2.14
DS::PMF	0.445	0.448	2.14
GOLD::ChemScore	0.441	0.452	2.15
SYBYL::D-Score	0.392	0.447	2.19
DS::Jain	0.316	0.346	2.24
GOLD::GoldScore	0.295	0.322	2.29
SYBYL::PMF-Score	0.268	0.273	2.29
SYBYL::F-Score	0.216	0.243	2.35

Bioinformatics 2010, 9:1169-1175. doi: 10.1093/bioinformatics/btq112

Generic Machine Learning Scoring Functions for Binding Affinity

- Other models including Support Vector Regression, Neural Networks, Random Forests accounted for intermolecular interactions and physio – chemical ligand properties and performed even better

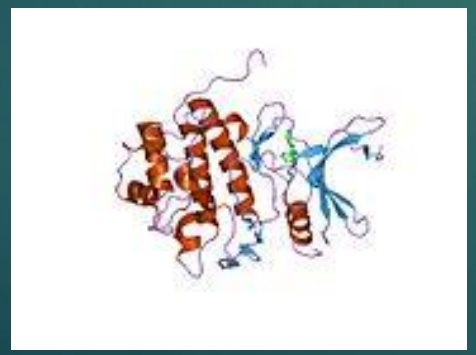


https://openi.nlm.nih.gov/detailedresult.php?img=PMC3877102_pone.0083922.g001&req=4

Family Specific Machine Learning Scoring Functions

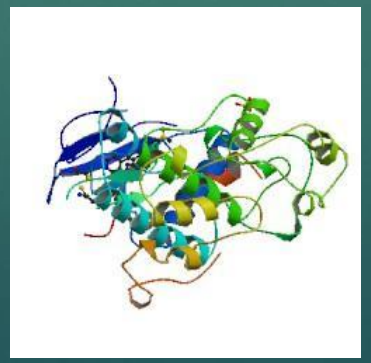
- Goal is to use Scoring Functions for specific drug targets
- Two ways to pick Scoring Functions for family specific targets from general Machine Learning Scoring Functions:
  - Pick best performing Scoring Functions on diverse set representing many target classes
  - Pick best performing Scoring Functions on test set of complexes of specific target class

CHK1



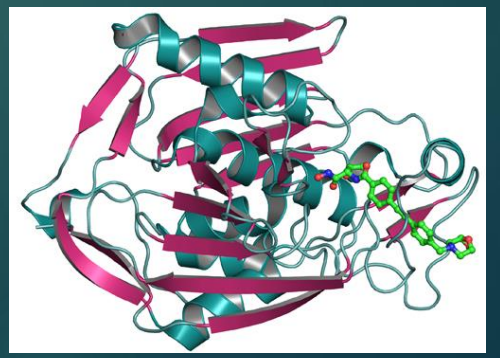
https://en.wikipedia.org/wiki/CHEK1

ERK2



http://www.phosphosite.org/proteinAction?id=832&showAllSites=true

LpxC

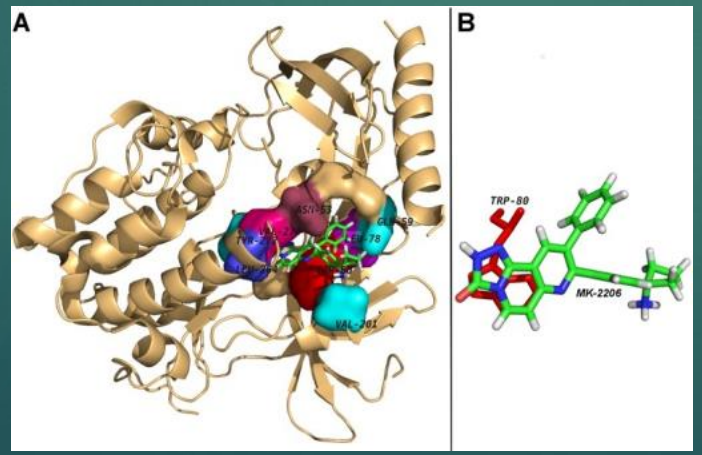


http://necat.chem.cornell.edu/Structures2/3NZK.html

Family Specific Machine Learning Scoring Functions

- Building family specific Scoring Functions – allows for more specific features to that target class
- Unsure if general or family specific Scoring Functions perform better for one target class because complexes from other target classes can contribute to performance

MD – SVR for Akt1 Inhibitors



https://openi.nlm.nih.gov/detailedresult.php?img=PMC4201482_pone.0109705.g001&req=4

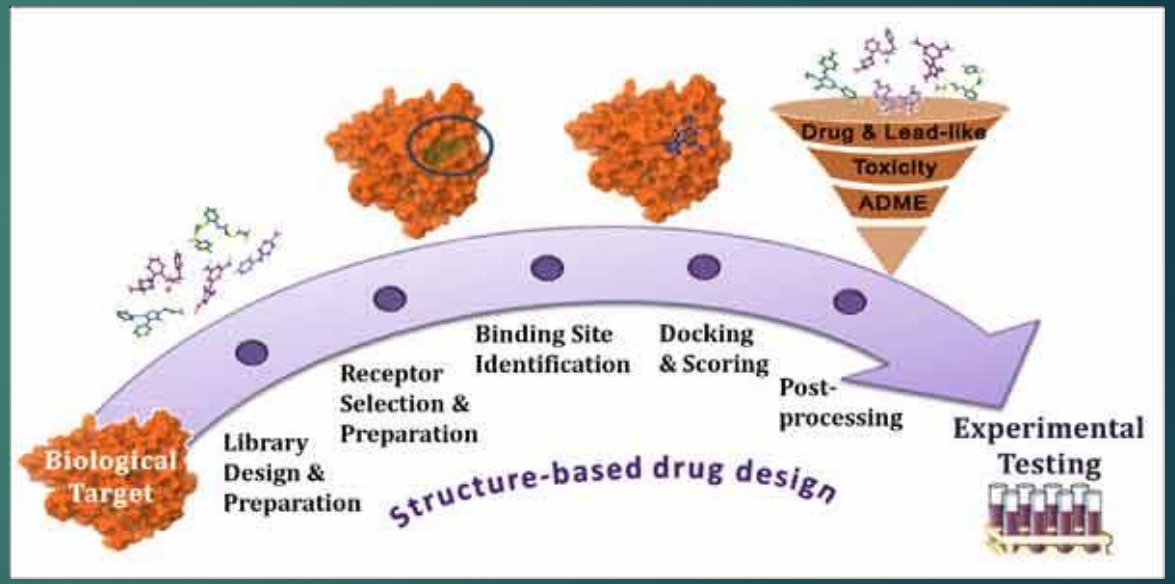
Machine Learning Scoring Functions for Virtual Screening

- Two types of Machine Learning Scoring Functions for virtual screening:
  - Regression based for ranking molecules – similar to binding affinity models
  - Classifiers for Virtual Screening – whether molecules will bind, true binders vs decoys
- Many models for classifiers including Random Forest, Support Vector Machine, Neural Network, Naïve Bayes

http://www.nature.com/nchembio/journal/v6/n5/full/nchembio.354.html?message-global=remove

Machine Learning Scoring Functions for Virtual Screening

- Important results from experiments
  - Importance of tailoring machine learning Scoring Functions to task (Binding Affinity or Virtual Screening)
  - Training higher number of actives and inactives improves performance



http://www.eurekaselect.com/124979/article

Emerging Applications of Machine Learning Scoring Functions

- Pose generation
- Molecular Recognition
- Drug lead optimization
- Protein – protein binding affinity Scoring Functions

Limitations

- General summary of machine learning scoring functions, not too much detail on specific models
- No mention of performance of pose generation prediction models and importance to docking compared with scoring
- Feature selection is key for improving performance and is very difficult to accurately choose features for models

A Whole New Scorer

Docking predicts binding strength
By using scoring functions
Classical scorers
Are bad so we use machine learning ones

Generic functions work
On many diverse complexes
Better results than classic functions
And improvements with train size

A Whole New Scorer
For predicting binding affinity
Using machine learning
Improves scoring
Applied to virtual screening

https://youtu.be/FSzpEE46PMY?t=21s

A Family Specific
Scoring function for drug targets
Can be from generic ones
Or family specific
With detailed features for that target class

Applied to virtual screening

Can be regression based
Or classifying binders
Models for molecule finders
Like naïve Bayes and neural nets

A Whole New Scorer
For predicting binding affinity
Using machine learning
Improves scoring
Applied to virtual screening for increased accuracy

http://aladdin.wikia.com/wiki/Genie

References

- Bioinformatics 2010, 9:1169-1175. doi: 10.1093/bioinformatics/btq112
- WIREs Comput Mol Sci 2015, 5:405–424. doi: 10.1002/wcms.1225
- http://www.intechopen.com/books/protein-engineering-technology-and-application/protein-protein-and-protein-ligand-docking
- http://archive.cnx.org/contents/4e7287b0-6c38-4829-abeb-3ae357bbf60f@10/protein-ligand-docking-including-flexible-receptor-flexible-ligand-docking
- https://openi.nlm.nih.gov/detailedresult.php?img=PMC3877102_pone.0083922.g001&req=4
- https://en.wikipedia.org/wiki/CHEK1
- http://www.phosphosite.org/proteinAction?id=832&showAllSites=true
- http://necat.chem.cornell.edu/Structures2/3NZK.html
- https://openi.nlm.nih.gov/detailedresult.php?img=PMC4201482_pone.0109705.g001&req=4
- http://www.nature.com/nchembio/journal/v6/n5/full/nchembio.354.html?message-global=remove
- http://www.eurekaselect.com/124979/article
- http://aladdin.wikia.com/wiki/Genie